ACOUSTIC METHODS IN PSYCHIATRY

by PETER F. OSTWALD

REPRINTED FROM
SCIENTIFIC AMERICAN
MARCH 1965
VOL. 212, NO. 3 PP. 82-91

PUBLISHED BY W. H. FREEMAN AND COMPANY 660 MARKET STREET, SAN FRANCISCO, CALIFORNIA 94104

 COPYRIGHT © 1965 BY SCIENTIFIC AMERICAN, INC. ALL RIGHTS RESERVED.

Acoustic Methods in Psychiatry

In which the speech of disturbed persons is analyzed not for content but for such attributes as loudness and timbre. The results indicate that this analysis can be useful in the description of mental state

by Peter F. Ostwald

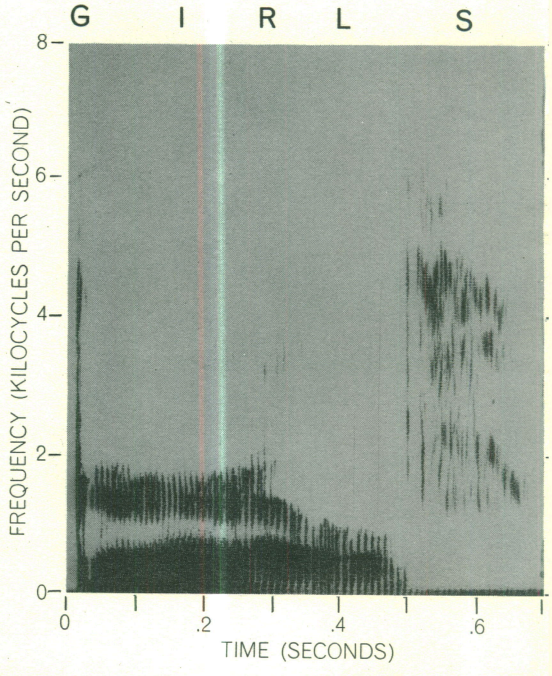
When someone calls on a physician, usually the first thing he does is talk about the symptoms that have led him to conclude he may be ill. The physician translates this description into a written history that, along with the results of a physical examination and any necessary X-ray or laboratory tests, helps to establish a working diagnosis and plan of treatment. The patient's oral testimony not only provides indirect clues about the illness but also on occasion offers direct acoustic evidence as to the pathological process that caused it. A tumor of the larynx or vocal cords, for example, imparts a hoarse quality to the voice. Brain damage due to a blood clot prevents the patient from thinking of certain words to describe his condition; hence his speech contains gaps, errors or peculiar word substitutions.

Clearly not all acoustic evidence of disease is obvious, and not all patients make an effort to reveal their problems. An emotionally disturbed patient may be unaware of his difficulties and inner conflicts, or he may try to hide them, fearing embarrassment, rejection or even punishment. Psychiatrists, who attach considerable diagnostic importance to the patient's history, must be particularly alert to acoustic manifestations of illness. When a patient begins to lose control over his emotions, he often betrays his distress by slips of the tongue, mispronounced words or changes in the quality of his voice. If the patient suffers from a disorganizing illness such as schizophrenia, he may alert the psychiatrist's "third ear" by rambling, repeating himself or using bizarre terminology or words he has coined himself. The speech of the schizophrenic patient may also have a peculiar blandness of expression called "flat affect."

The written history, even if it includes actual quotations or a letter in which the patient describes his disturbance, remains a soundless document. On occasion this is an advantage: some abnormalities in a patient's verbalized thoughts become clearer when they are set down on paper. But the written history also has a major shortcoming, apparent to anyone who has winced at a novelist's reproduction of a Southern drawl. Speech has nonlexical features that are almost impossible to capture on paper, such as intonation, tempo, loudness and timbre. Such features are affected by illness. A person suffering from multiple sclerosis, for example, develops a "scanning" type of speech that can only be recognized when it is heard. Psychoneurotic patients who stutter produce sounds that defy translation into words.

The inherent limitation of the written history has stimulated physicians to search for more precise methods of describing the sounds made by their patients. As long ago as 1827 James Rush of Philadelphia published a book entitled *The Philosophy of the Human Voice,* in which he used a kind of musical notation to depict the nonlexical aspects of speech. In recent years Paul J. Moses, an eminent laryngologist at the Stanford University School of Medicine, has developed a method of analyzing human vocalization in terms of the "five *R*'s": respiration, range, register, resonance and rhythm. My own search for ways to study the speech of psychiatric patients has led me to try a radically different method of describing voice patterns. My colleagues and I at the University of California School of Medicine in San Francisco have borrowed the techniques developed in acoustic laboratories for precise analysis of sounds and have put them to clinical use in analyzing speech.

With the help of George Wilson, an acoustical engineer at the University of California at Berkeley, we undertook to determine if one could characterize the sound produced by a psychiatric patient in purely physical terms. We set up a tape recorder in a room where a psychoneurotic woman with dangerous trends toward depression and overeating was to be interviewed. During a series of interviews we recorded her speech through a microphone 10 inches from her mouth and in full view. After selecting four-second segments of tape containing her most voluble utterances (which struck us as having a childish tone), we spliced the segments into loops of continuous tape and ran the loops



NORMAL SPEECH PICTURE, made by sound spectrograph from a healthy man's pronunciation of the word "girls," shows distinctive pattern associated with each letter.

through a sound analyzer. This device divided the constituent frequencies of the sound into bands either one octave wide or half an octave wide. The amount of energy in each band was then plotted against frequency to obtain a curve that represented a characteristic sound pattern of the speech being analyzed.

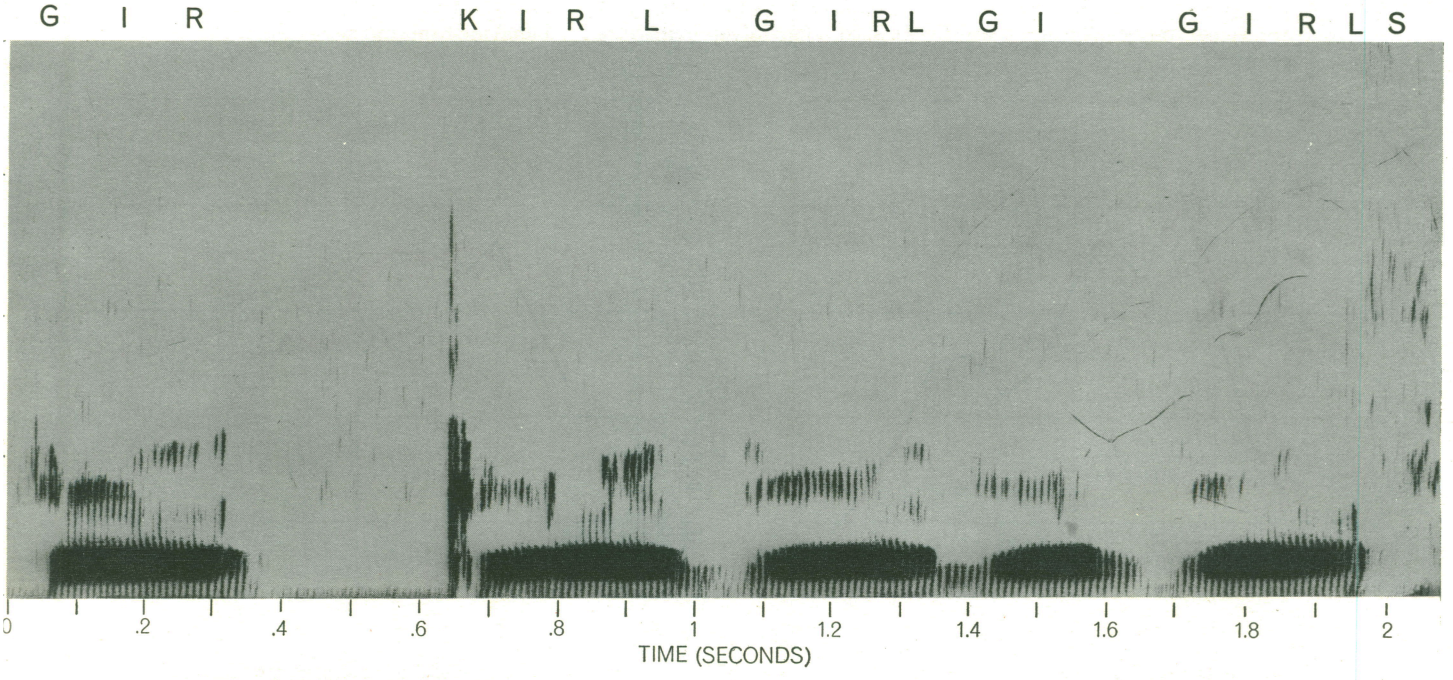
When human beings speak, most of the sound energy they produce lies at frequencies between 100 and 10,000 cycles per second, the pattern of frequencies varying with the age and sex of the speaker and the words spoken. The sound analyzer showed the speech of our first patient to have a fundamental focus of energy at 250 cycles per second (close to middle *C* on the musical scale) and a focus of resonance energy at 500 cycles per second—one octave higher. After repeatedly detecting the same distribution of frequencies in her voice and the voices of patients who sounded much the same, we dubbed this particular pattern the "sharp" voice. Later we obtained an almost identical acoustic pattern from the speech of a 10-year-old boy, which supported our impression that the sharp voice has a childish quality.

With our technique of sound analysis we have also developed quantitative definitions for such vocal stereotypes as "flat," "hollow" and "robust" [*see illustrations at bottom of next two pages*]. The flat voice—which can be equated with the monotonous, low-pitched delivery of patients in states of depression—yielded a pattern of sound energy spread quite evenly over a broad band of the frequency spectrum. The hollow voice—which is common to extremely retarded, inhibited or repressed patients and victims of certain types of brain disease—can be described as an exaggerated kind of flat voice; it has one prominent energy focus representing the fundamental tone of the voice and has little resonance energy. The robust voice—which is characteristic of patients who speak in an oratorical, overly emphatic way, often with some pomposity—yielded a striking pattern of relatively large amounts of sound energy concentrated at the center of the frequency spectrum in a symmetrical manner.

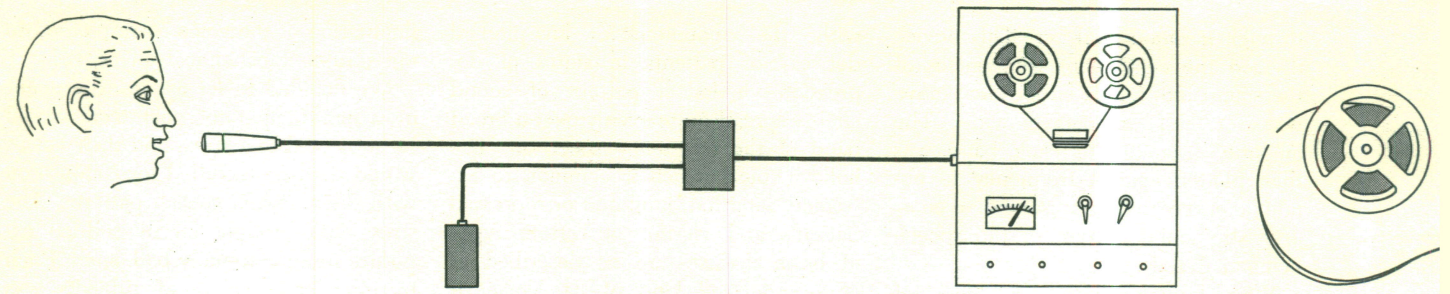
The use of acoustic methods to describe the overall quality of a patient's voice suggested that one might be able to use such methods to detect the detailed variations that emotional stress causes in the voice. Encouraging support came from some work done by Leonard Rubenstein of McGill University. He had asked experimental subjects to hum a certain tone during the day and then to hum what they thought was the same tone at night. He found that the frequency pattern of the hummed tone changed regularly from day to night, and that this diurnal fluctuation of the voice corresponded to measurable biochemical changes.

We decided to set up an experiment in which nonpatient volunteers would be subjected to a stress and their voices would then be recorded for comparison with their normal speech patterns. The stress was brought on by sudden exposure to an extremely bad smell. Each member of a group of subjects was asked to come into a room, sit down in front of a microphone and read "Joe took father's shoe bench out," a test sentence chosen for its variety of sounds. After the sentence had been repeated three times the subject was instructed to pick up a glass of what appeared to be water but in reality was a concentrated solution of ammonia. He was told to take a deep breath from the glass and repeat the test sentence into the microphone. Members of a second group of subjects were asked first to smell the ammonia and later on to record their normal voice; this was done to remove the possible effects of the sequence of procedures from the final results of the experiment. When we examined the half-octave-band analyses for the subjects, we found that the voices recorded after olfactory stress contained strikingly less energy than normal voices in the half-octave band centered at 500 cycles per second.

Having determined that the testing technique would work, we undertook a similar study of a group of psychiatric



ABNORMAL SPEECH PICTURE of the efforts of a stutterer to say the word "girls" was made by sound spectrograph. A weak opening "g" is followed by adequate vowel formants (*resonance energy rising in streaks*) but the final "ls" is missing. A second start results in a sound resembling a "k." After two more false starts the word is completed. A spectrogram charts frequency in kilocycles per second (*vertical axis*) against time in seconds (*horizontal axis*). The author added an alphabetized transcription.



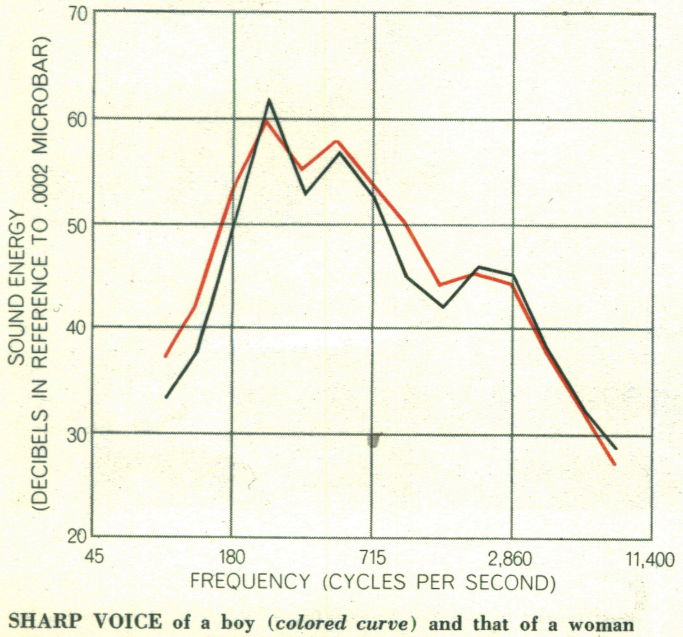
METHOD OF ANALYZING SPEECH used by the author in his earliest acoustic studies involved recording a patient's speech, splicing samples of it into continuous loops of tape and playing these loops backward through a tape recorder. A sound analyzer

patients. Because we did not want to expose the patients to any kind of additional stress we dispensed with the bad-smell stimulus and compared the sounds they produced before and after psychiatric treatment. Each patient was interviewed as soon as possible after being admitted to the treatment ward in an acute state of emotional disturbance. During the interview, which was tape-recorded, we asked the patient to read aloud "Joe took father's shoe bench out." After about three weeks of treatment by the ward physicians the patients were again interviewed and asked to repeat the test sentence.

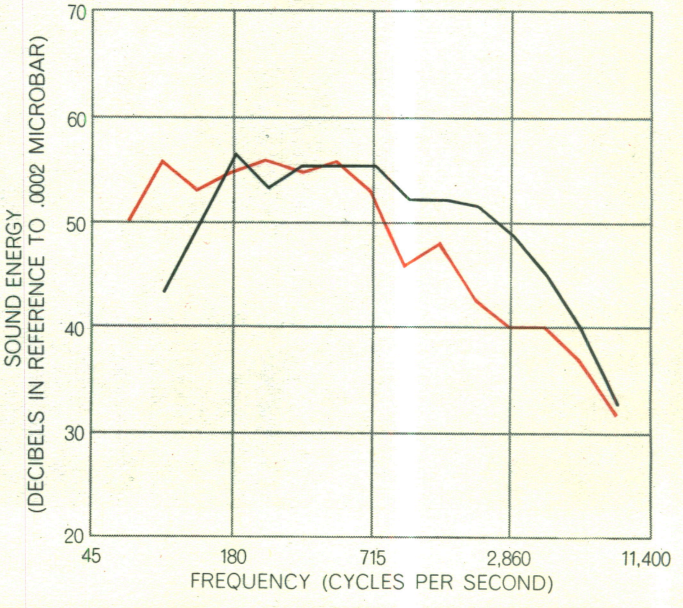
The "reading voice" adopted by a patient uttering a test sentence does not provide a sample of spontaneous speech. To compensate for this we compiled two other samples of a patient's voice during each recorded interview, one consisting of high-pitched segments of sound and the other of low-pitched segments. For purposes of analysis we then considered not the voice of a patient but his voices: the high voice, the low voice and the reading voice.

One man we studied in this way entered the hospital in a state of severe depression resulting from manic-depressive psychosis; his voices were uniformly flat. After a series of electroconvulsive treatments he spoke in such a way that his high voice assumed a robust pattern characteristic of exuberance and overconfidence. His reading voice also became more robust, but his low voice retained the basically flat configuration observed before treatment. Taking into account only this acoustic evidence, we surmised that the patient had undergone a profound clinical change, tending outwardly toward aggressive, extroverted behavior but inwardly toward a sense of sadness reflecting the depression that had necessitated his coming to the hospital. This surmise was verified by the physicians and nurses in the ward who had been following the patient's daily course in detail.

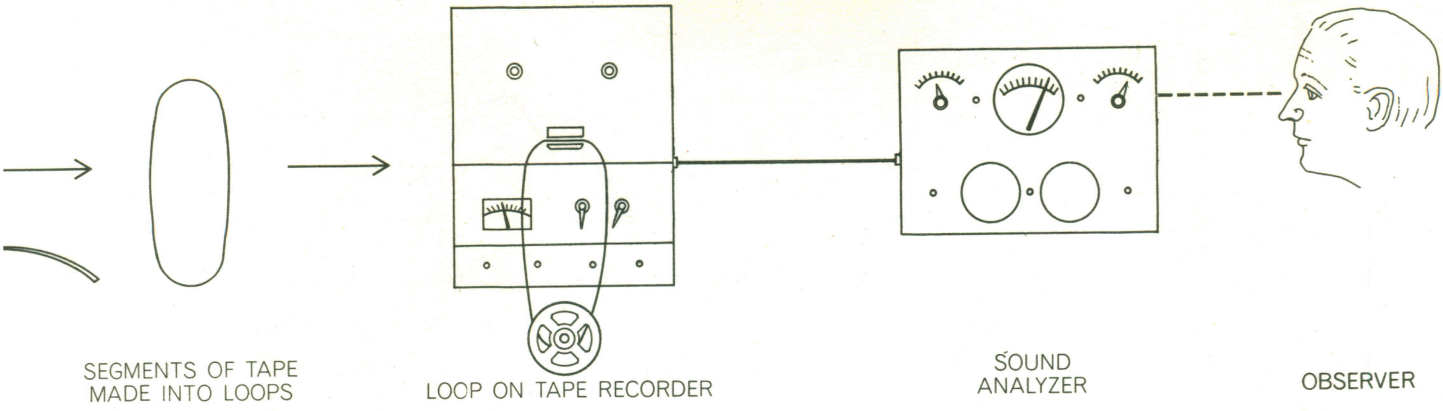
Another case we studied was that of a woman admitted for treatment of a type of schizophrenia that involved bizarre behavior and almost total social withdrawal. Before she had been treated her voice was audible but sounded subdued, a condition we traced acoustically to a sharp low voice, a sharp reading voice and a flat high voice. After she had been treated with a phenothiazine drug known to ease acute schizophrenic reactions, there was a greater difference between the voices, indicating greater variability in her behavior from moment to moment. Her



SHARP VOICE of a boy (*colored curve*) and that of a woman (*black curve*) are analyzed. Each has peaks a single octave apart. Level of sound energy (*vertical axis*) was measured along half-octave bands. Frequency (*horizontal axis*) is in cycles per second.



FLAT VOICE, subjectively associated with monotonous speech, is defined by acoustic analysis as a pattern of sound energy spread quite evenly over the frequency spectrum with few resonance peaks. Curves show the speech of a man (*color*) and a woman (*black*).



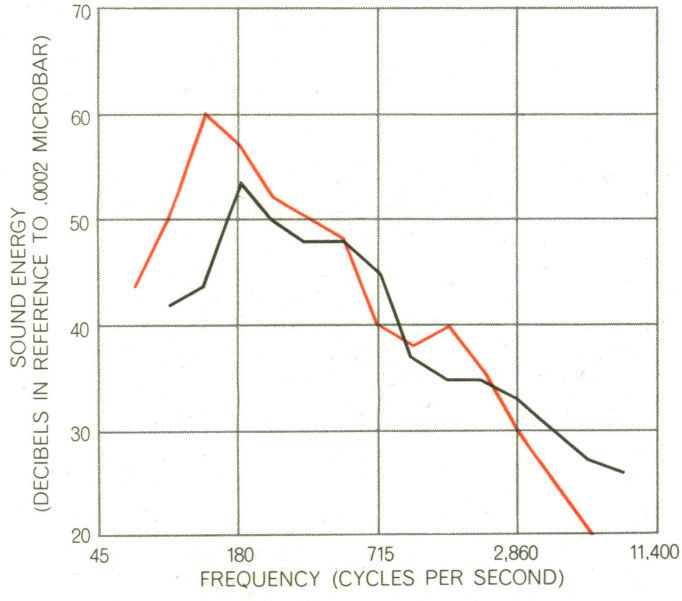
would then indicate the level of sound pressure (energy) existing within consecutive half-octave and octave bands on the frequency spectrum. The needle of the sound analyzer was observed as it fluctuated so that data for graphs could subsequently be compiled.

voices after treatment also showed an increase in the amount of total energy produced, suggesting that she had become more assertive and outgoing. There also appeared a tendency toward the flat-voice or hollow-voice pattern, implying that her feelings of depression had increased or that the drug had exerted some inhibiting effect on her vocal behavior.

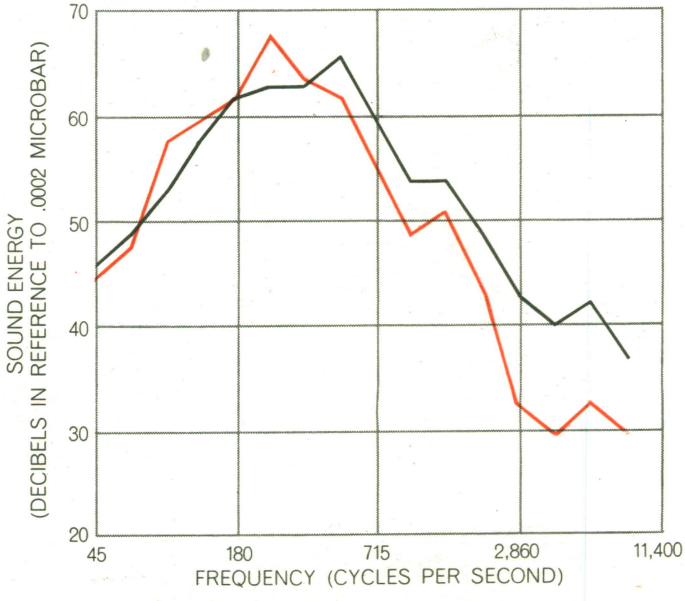
A third patient whose voices we analyzed was a woman admitted to the hospital in an acute state of agitation and excitement diagnosed as a hypochondriacal panic reaction. Both her high voice and her reading voice showed the robust pattern associated with an overly dramatic manner, whereas her low voice was flat, indicating depression. When the patient's voices were analyzed after treatment, they had become more alike; the robust high voice had come down and the formerly flat low voice was almost up to the robust reading voice. The regime of treatment in this case had consisted of careful ward management coupled with intensive group therapy and individual psychotherapy. Once the patient's attention was diverted from her physical condition she became a more controlled, easygoing individual, although she was still given to a theatrical manner, as indicated by her predominantly robust voice.

Since we had obtained identical test sentences from all the patients before and after treatment, it was possible for us to say something about how they changed as a group. We used the results of acoustic analyses of the reading voices to test for any statistically significant associations with such clinical elements as diagnosis and form of treatment. One association seemed readily apparent: the use of electroconvulsive treatment consistently led to increased sound-energy levels over the entire acoustic spectrum, particularly at 500 cycles per second.

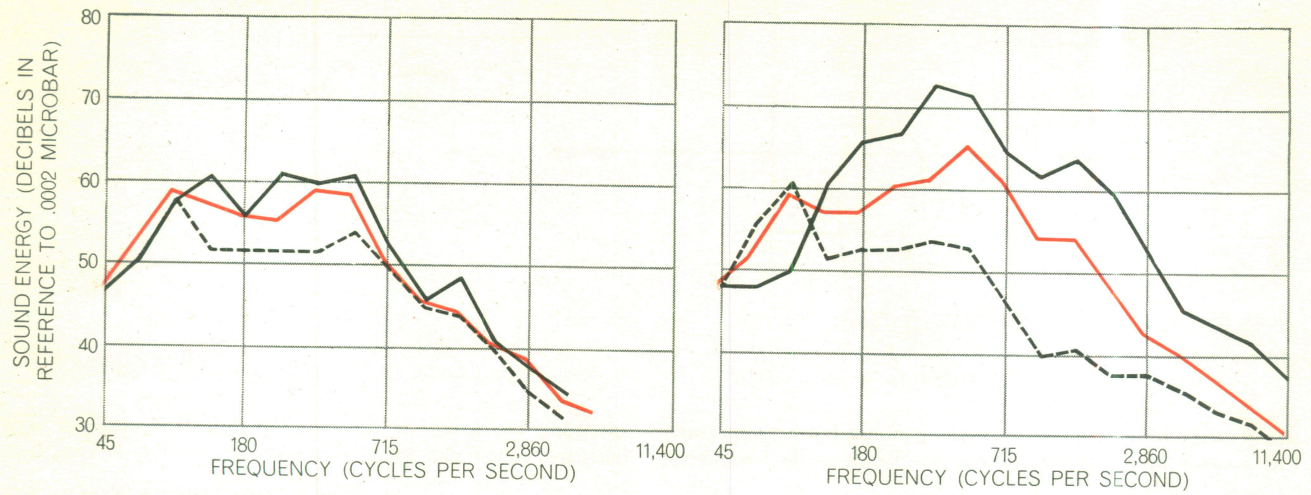
The persistent way in which energy changes centering at 500 cycles per second showed up in our studies turned out to have some biological significance. It is a fundamental fact of auditory perception that the human ear is most sensitive to tonal stimuli above 500 cycles per second. Almost all sounds calculated to arouse human beings—whistles, cries, sirens, screams—contain a large amount of sound energy above 500 cycles per second. On the other hand, the fundamental tone of the human voice, which is located at lower frequency levels, is so dispensable in



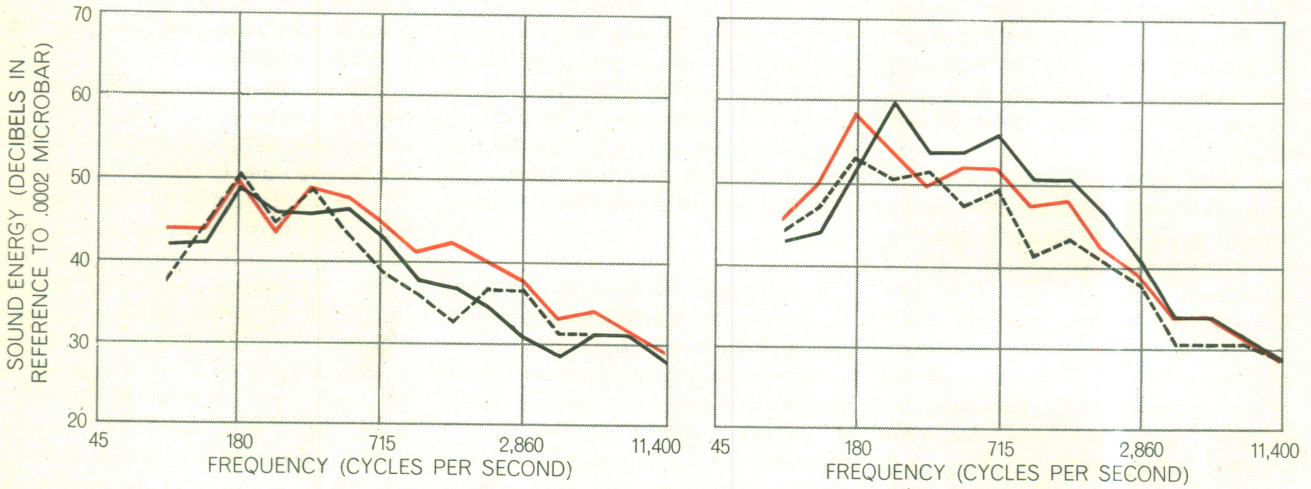
HOLLOW VOICE is a vocal category characterized by a single-peaked curve. The peak indicates that only the fundamental tone of the voice is carrying much sound energy. The level of energy, or sound pressure, is in decibels in reference to .0002 microbar.



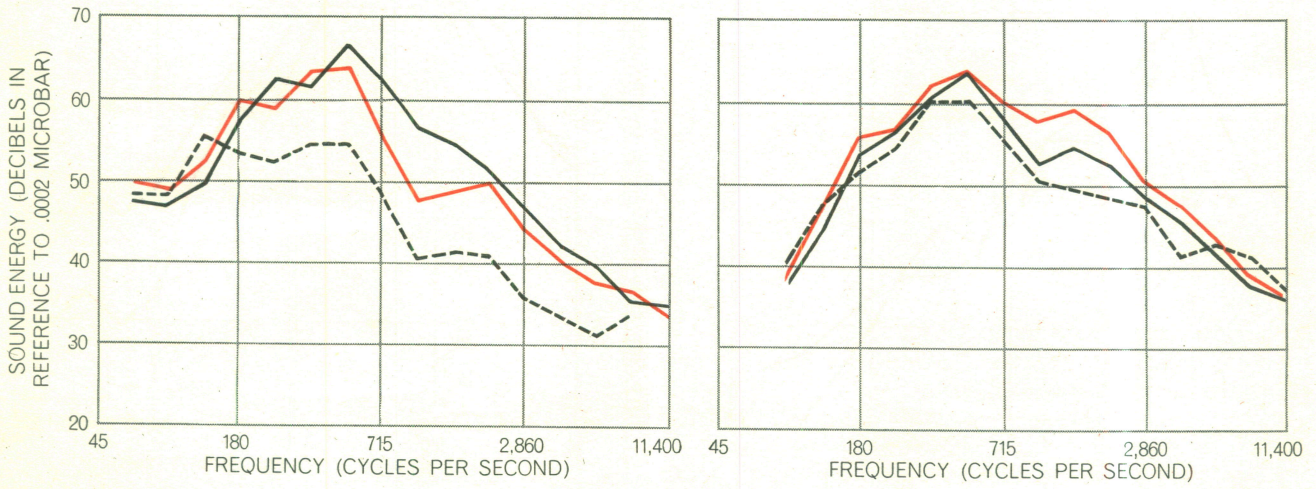
ROBUST VOICE, commonly identified with oratorical speech, makes use of more sound energy than other vocal categories. In samples from a man (*color*) and woman (*black*) sound energy is spread somewhat symmetrically over the frequency spectrum.



DEPRESSED PATIENT studied by author changed speech pattern in hospital. Before treatment his three voices seemed flat (*curves at left*). After therapy his high and reading voices took on a robust pattern, whereas his low voice remained flat (*curves at right*), implying slight recovery. Curves are keyed to voice types.



SCHIZOPHRENIC PATIENT sounded subdued when admitted to hospital. Her voices lacked energy and were uniform in pattern (*curves at left*). After treatment her voices acquired energy and variety (*curves at right*). Persistence of flat and hollow patterns, however, indicated that her vocal behavior was inhibited.



FRIGHTENED PATIENT entered hospital in a disorganized mental state. Her panic was reflected in robust high and reading voices; her feelings of depression were suggested by a flat low voice (*curves at left*). After therapy her voices gained similarity to each other (*curves at right*), implying better personal control.

ordinary conversation that telephone companies do not even bother to transmit it. We advance the hypothesis that speakers increase the amount of sound they emit at higher frequencies when they want to be listened to and reduce their high-frequency output when they prefer to be ignored.

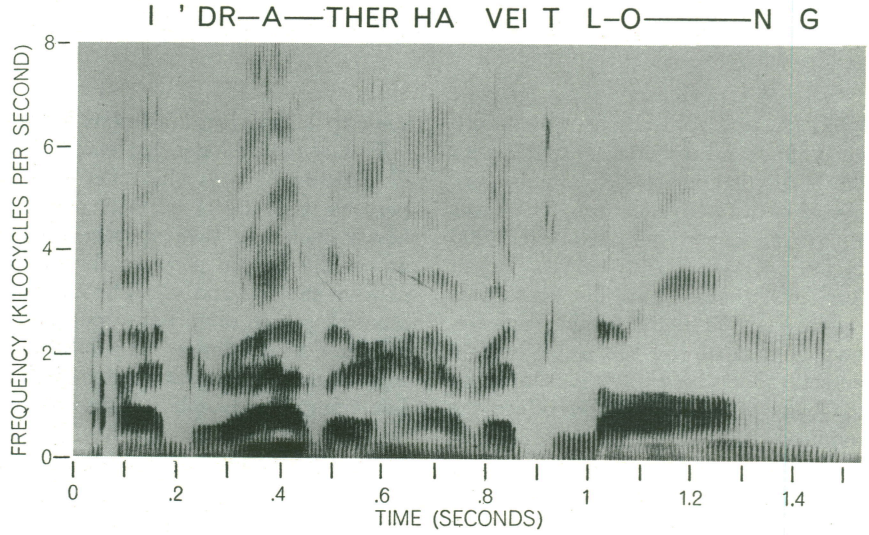
We departed somewhat from our main line of study to examine in detail a human sound specifically designed to attract attention: the crying of a baby. Analysis showed that babies' cries invariably have a fundamental tone located around 500 cycles per second and resonance bands appearing at 1,000, 2,000 and 4,000 cycles per second. Nonetheless, the individuality that leads mothers to use such diverse characterizations as "lusty," "strong," "loud," "feeble," "weak" or "whiny" in describing a given baby's cry was not reflected by the half-octave-band method of acoustic analysis. Eventually our interest in such detail led us to abandon the sound-analyzer method for a more sophisticated technique: sound spectrography.

The sound spectrograph is a device that analyzes sound into bands of frequencies and converts it into electrical impulses that, when they are recorded on a special paper, provide a vivid picture of the particular sound under observation. Workers in acoustic phonetics describe consonant sounds as "noise" and vowel sounds as "formants": discrete bands of resonance energy that vary according to the vowel being uttered. A sound spectrogram—the picture made by the spectrograph—presents vowel formants as heavy, dark bands and consonant noise as unevenly scattered gray spots. Normal speech has characteristic formant and noise features; it is therefore easy to detect abnormalities in the speech of psychiatric patients by looking at their sound spectrograms. The spectrogram also makes it possible to analyze speech disturbances in considerable detail.

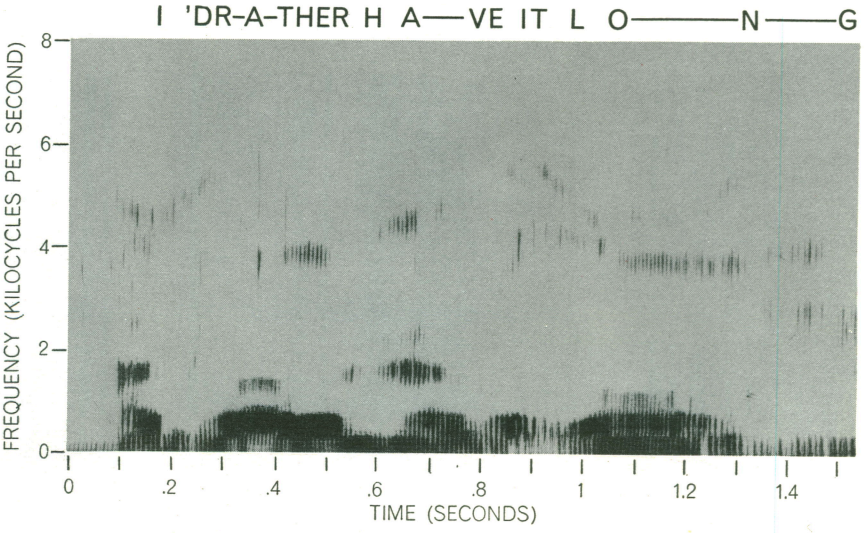
A disturbance that readily lends itself to analysis by spectrogram is stuttering. Regularly associated with other behavior disturbances, this condition occurs more frequently among males than among females. The voice of a young man who stuttered and was under treatment for a distressing emotional conflict was recorded by sound spectrograph as he attempted to pronounce the word "girls" [*see illustration on page 3*]. He began with a weakly articulated "g," and although he enunciated the following vowel and the consonant "r," he could not manage the terminal consonant cluster "ls" to complete the word. Instead the patient fell silent for about three-tenths of a second and started over again. He then made too forceful an attack on the "g," giving it more the properties of "k." Again the vowel followed, but as before he could not complete the word. He started twice more without success. Only on the fifth try was he able to produce the consonant noise needed to complete the word.

A speech disorder more elusive and difficult to record, and one that has interested theologians as well as psychiatrists, is glossolalia, or "talking in tongues." Victims of this disturbance are quite capable of producing normal speech, but in states of intense emotional excitement they make peculiar and incomprehensible sounds. Spectrograms of the sounds show that they have such properties of normal speech as formants, noises and brief separations due to the articulation of speechlike fragments. The sounds are uttered, however, without the logical structure necessary for intelligibility [*see top illustration on next page*].

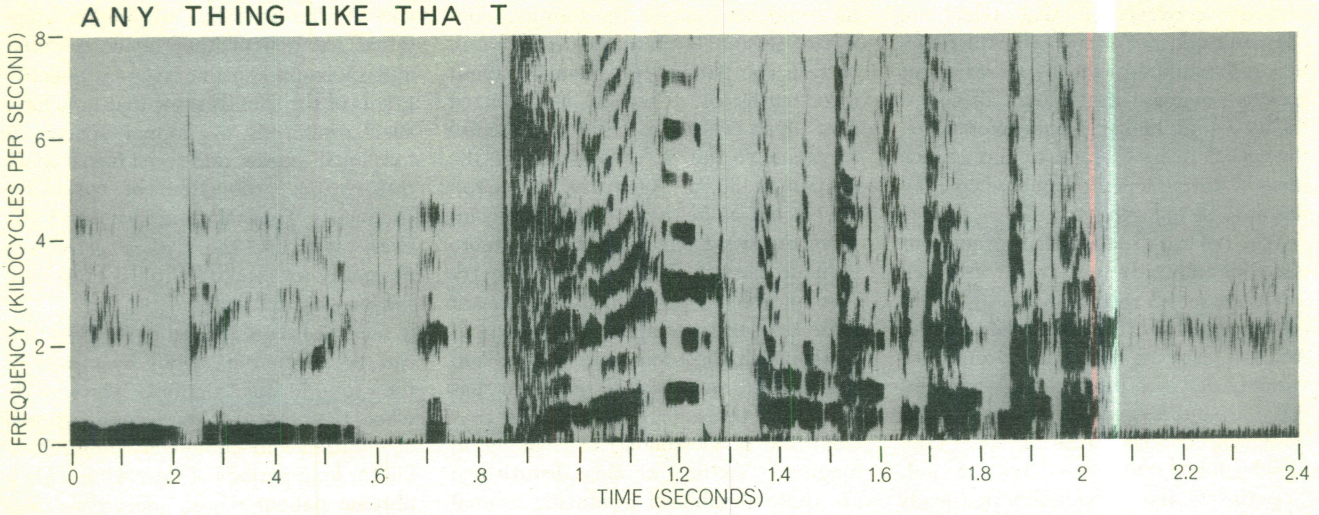
Schizophrenia is another disorder that can be studied in detail by means of the sound spectrograph, particularly when that disorder renders the speech of children and adolescents unintelligible for long periods of time. One schizophrenic patient whose speech we ana-



SPECTROGRAM OF INTELLIGIBLE SPEECH makes the vowel formants readily apparent as resonance energy bands rising throughout the frequency spectrum (*vertical axis*). The spectrogram also depicts a clearly articulated stop consonant (the "t" of the word "it").



SPECTROGRAM OF UNCLEAR SPEECH represents the attempt of a schizophrenic boy to say "I'd rather have it long." The spectrogram lacks distinctive vowel formants, precise consonants and significant output in the part of the spectrum above 1,700 cycles per second.

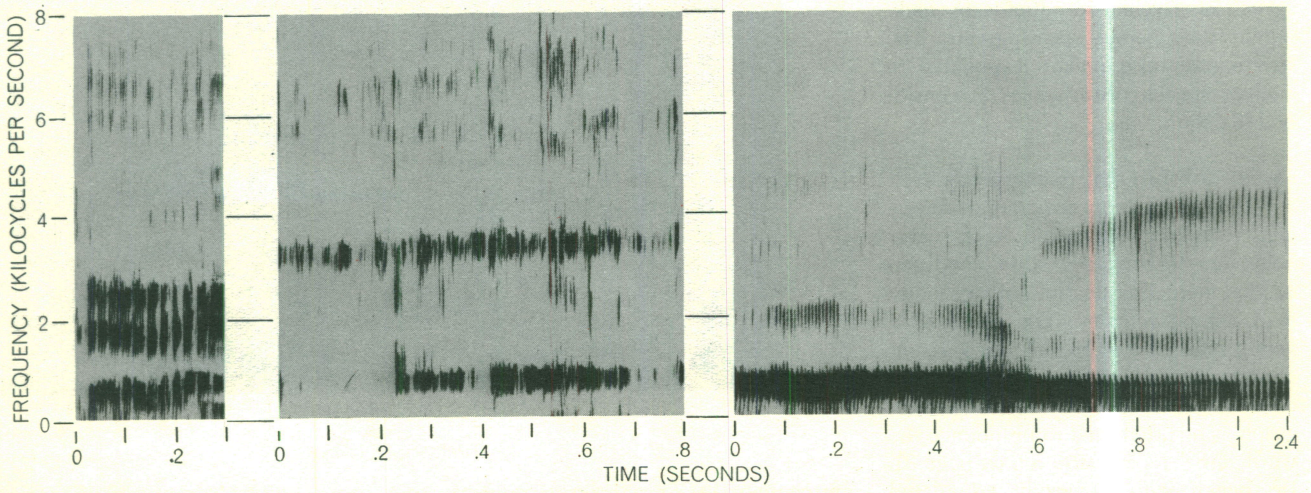


SPECTROGRAM OF GLOSSOLALIA, or "talking in tongues," shows normal speech ended by a burst of noise at .8 second on the horizontal axis. Ensuing is a series of rising formants with a resonance pattern exceeding 8,000 cycles per second (*vertical axis*). A break at 1.3 seconds is followed by a drop in pitch and some short bursts with formants and noise resembling true speech sounds.

lyzed spectrographically kept his mouth open almost all the time he spoke, with the result that his speech sounds were devoid of distinctive features. There was almost no acoustic output above 1,000 cycles per second, his vowels lacked formant structure and his consonants produced little noise in the high-frequency part of the spectrum. Such a defect cannot be explained simply on the basis of inability to reproduce the required phonetic properties of speech. As we observed the patient during various emotional states it became apparent that he could in fact vocalize energetically, but in doing so he often made sounds that completely lost the character of human speech and turned into something more reminiscent of the screeches, rasps, squawks or growls made by animals [*see illustration below*].

Part of the problem that emotionally ill patients have is that they cannot perceive the effects of their own behavior—including their speech—as well as healthy people perceive the effects of what they do and say. The emotionally sick person often has a long-standing sense of inferiority; his painful feelings of shame or guilt tend to isolate him from other people or cause him to deal with them as though they did not exist. As a result he gets little or no self-corrective feedback from the social environment and his behavior becomes increasingly bizarre, leading to the diagnosis of mental illness. For dealing with such emotionally sick people the new techniques of acoustic analysis can be combined with clinical and experimental methods to study behavior. John A. Starkweather and William A. Hargreaves of the Langley Porter Neuropsychiatric Institute in San Francisco have developed an ingenious way of studying voice changes during psychotherapeutic interviews. Using a digital computer they correlate the sounds made by the patient with variations in his mood from moment to moment.

In the summer of 1963 I witnessed a particularly imaginative combination of acoustic analysis and personal interaction in the treatment of a withdrawn, schizophrenic boy of eight. The treatment was undertaken by Mary O'Keeffe, a student at the University of California School of Medicine who was working on a summer research fellowship. As a



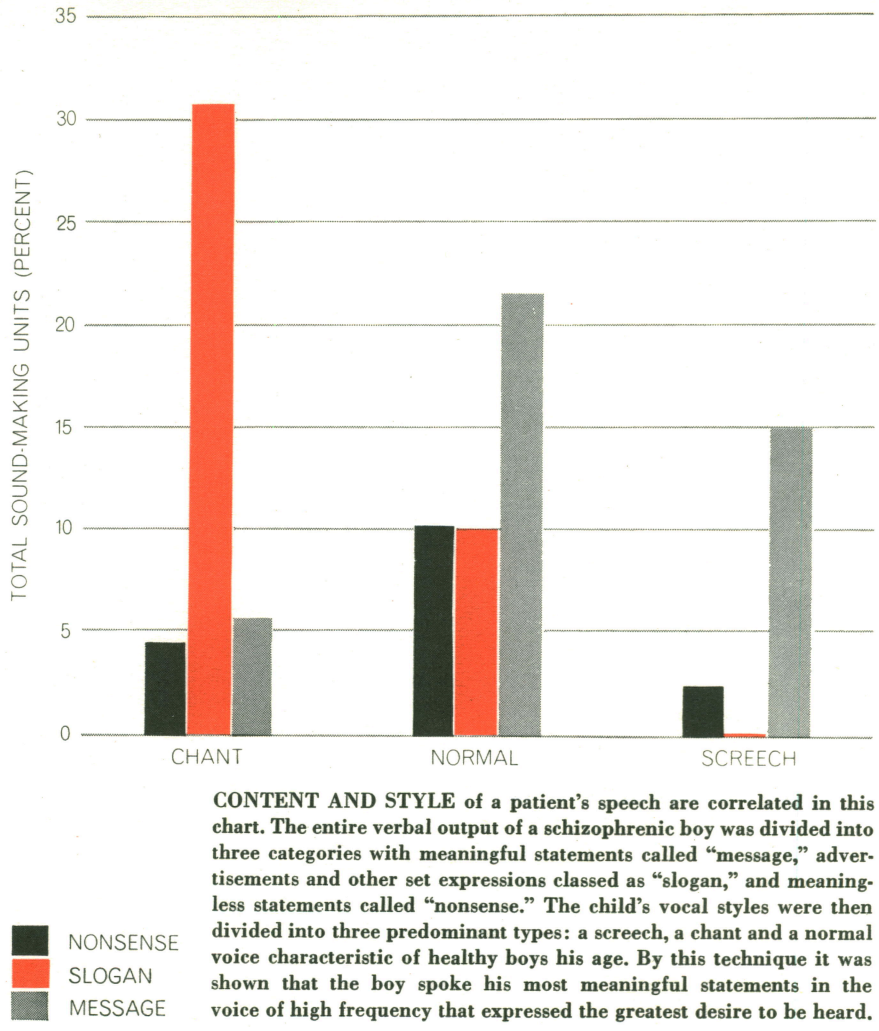
SPECTROGRAMS OF ANIMAL-LIKE SOUNDS were made during the outbursts of a disturbed boy. The sound depicted at left was described by the author as a rasp; the one in center, as a froglike squeak-squawk; the one at right, as a growl sounding like "aw-er."

psychiatrist at the School of Medicine I served as her consultant.

Miss O'Keeffe's patient was restless and frequently unmanageable. He would dart about the room or freeze into an immobile posture for no apparent reason. He grimaced and moved his hands in a strangely stereotyped manner. Often his speech seemed akin to the babblings and mouthings of babies. There were other times when the boy produced recognizable fragments of words and sentences, disconnected and without grammatical order but revealing an unusually large vocabulary. Much of this vocabulary consisted of words he had heard but could not use for meaningful communication. Miss O'Keeffe spent many hours observing the boy in the ward and in the playground. She also tried to speak with him, and she tape-recorded these sessions whenever possible.

Analysis of the tape recordings revealed that the content of the patient's speech fell into three distinct categories. The first category—message content—consisted of words and phrases that seemed relevant to the situation confronting the patient and conveyed information in a semantic sense, such as "I want to go out." The second category—slogan content—included words from songs, rhymes, television advertisements and other incidental verbal material that the patient would memorize and compulsively repeat. Slogan content seemed related to his internal problems, fears and concerns and never to the actual situation in which he found himself. The third category—nonsense content—contained sounds in which it was impossible to discern any words whatsoever, and it also included pseudowords superficially resembling speech.

When the patient's acoustic productions were separated in terms of content, the vocal characteristics of these productions were studied and three different voice types became distinguishable. In addition to the "normal" voice characteristic of a healthy child there was a "screech" of high, unvarying frequency and much noise, and a "chant" containing gliding sounds and imparting a singsong effect. An attempt to relate these three different voice types to the differences in speech content indicated that the screech almost exclusively accompanied message content [*see illustration at right*]. That is, at moments when the patient sounded most agitated, he produced meaningful statements almost exclusively. Miss O'Keeffe and I concluded that the patient reserved his verbally meaningful speech for times when he indicated, by screeching, the urgency of being heard.



CONTENT AND STYLE of a patient's speech are correlated in this chart. The entire verbal output of a schizophrenic boy was divided into three categories with meaningful statements called "message," advertisements and other set expressions classed as "slogan," and meaningless statements called "nonsense." The child's vocal styles were then divided into three predominant types: a screech, a chant and a normal voice characteristic of healthy boys his age. By this technique it was shown that the boy spoke his most meaningful statements in the voice of high frequency that expressed the greatest desire to be heard.

Had it happened that people had paid no attention to him when he spoke more normally? Such a question can only be answered by studying the developmental background and behavior pattern of a patient through extensive interviews and direct observation. From an acoustic study we were able to learn that while using his normal voice the patient spent as much time emitting nonsense or slogans as he spent producing message content. This almost "fifty-fifty" split between meaningful and meaningless speech in his normal voice represented a characteristically ambivalent quality in his behavior. Whenever the patient spoke in a normal voice, he in effect both gave the listener information and took it away. Furthermore, he spent as much time chanting as he did using his normal voice, and the chanting was almost completely devoid of message content.

The behavior of schizophrenic children can be explained in terms of a number of unfavorable factors—some genetic, some physiological and some environmental—combining to interfere with the child's ability to learn language and to use speech for social communication. We have studied patients who could not produce the distinctive features of vowels and consonants. From such studies of people with brain damage as have been conducted by the Soviet neurologist A. R. Luria it seems likely that the information needed for normal speech is stored in the temporal zone of the cerebral cortex. If a patient who has suffered brain damage cannot recall or reproduce the distinctive features of the speech sounds, his utterances remain incomprehensible to others. Something much like this seems to happen in certain schizophrenic patients. We have reason to believe that further application of acoustic techniques to clinical phenomena in psychiatry will make possible precise descriptions of the communication disturbances associated with mental illnesses. One hopes that the information gathered by acoustic techniques will also provide new leads to effective treatment.

The Author

PETER F. OSTWALD is associate professor of psychiatry at the University of California School of Medicine in San Francisco, from which he was graduated in 1950. He also did his undergraduate work at the university. From 1951 to 1956 he was a resident in psychiatry at New York Hospital, returning then to the University of California School of Medicine. Ostwald reports that a childhood interest "in communication with sounds led to the study of music and to violin-playing" and that "during adolescence I became interested in psychoanalysis and decided to go into medicine." In addition to his teaching and research he maintains a private practice of psychiatry. He writes: "My interest is in psychotherapy as a way to explore and modify emotional problems, and I am curious to find out how scientific techniques can be developed and applied to the study of the doctor-patient relationship."

Bibliography

DISORDERS OF COMMUNICATION: PROCEEDINGS OF THE ASSOCIATION FOR RESEARCH IN NERVOUS AND MENTAL DISEASE. Edited by David M. Rioch and Edwin A. Weinstein. The Williams & Wilkins Company, 1964.

SOUNDMAKING: THE ACOUSTIC COMMUNICATION OF EMOTION. Peter F. Ostwald. Charles C Thomas, Publisher, 1963.

THERAPEUTIC COMMUNICATION. Jurgen Ruesch. W. W. Norton & Company, Inc., 1961.

THE VOICE OF NEUROSIS. P. Moses. Grune & Stratton, Inc., 1954.